AD/A-004 823

MINIMAX PARAMETER ESTIMATION UNDER GENERALIZED QUADRATIC LOSS WITH A COMPACT PARAMETER SPACE

Rafael Antonio, et al

Illinois University

Prepared for:

Joint Services Electronics Program

July 1974

DISTRIBUTED BY:

NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD/A-004823
4. TITLE (and Subtitle) MINIMAX PARAMETER ESTIMATION UNDER GENERALIZED QUADRATIC LOSS WITH A COMPACT PARAMETER SPACE		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER R-655; UILU-ENG 74-2221
7. AUTHOR(s) Rafael Antonio Padilla Lovera		8. CONTRACT OR GRANT NUMBER(s) DAAB-07-72-C-0259
9. PERFORMING ORGANIZATION NAME AND ADDRESS Coordinated Science Laboratory University of Illinois Urbana, Illinois 61801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Joint Services Clectronics Program Fort Monmouth, New Jersey		12. REPORT DATE July, 1974
		13. NUMBER OF PAGES 27
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Minimax Parameter Estimation

Quadratic Loss

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The subject matter of this work is minimax parameter estimation. In the present chapter we will introduce some basic concepts and definitions from the theory of parameter estimation and through these arrive at the results and techniques which we will use to search for minimax solutions to parameter estimation problems.

PRICES SUBJECT TO CHANGE

DD FORM 1 JAN 73 1473 EDI

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

MINIMAX PARAMETER ESTIMATION UNDER GENERALIZED QUADRATIC LOSS
WITH A COMPACT PARAMETER SPACE

by

RAFAEL ANTONIO PADILLA LOVERA

This work was supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAB-07-72-C-0259.

MINIMAX PARAMETER ESTIMATION UNDER GENERALIZED QUADRATIC LOSS
WITH A COMPACT PARAMETER SPACE

BY

RAFAEL ANTONIO PADILLA LOVERA

Ing. Elec., Universidad Central de Venezuela, 1972

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1974

Urbana, Illinois

ACKNOWLEDGMENTS

The author expresses his sincere gratitude to his advisor, Professor Max Mintz, for the encouragement and enthusiastic help he gave during the preparation of this thesis and for the honor of being one of his graduate students. Thanks are also due to Mrs. Rose Harris for her excellent typing of the manuscript.

The financial support provided by the UCV-UNESCO-VEN-31 Fellowship Program and the assistance of the Coordinated Science Laboratory of the University of Illinois is gratefully acknowledged.

TABLE OF CONTENTS

1. PRELIMINARY CONCEPTS

1.1. Introduction

The subject matter of this work is minimax parameter estimation. In the present chapter we will introduce some basic concepts and definitions from the theory of parameter estimation and through these arrive at the results and techniques which we will use to search for minimax solutions to parameter estimation problems.

1.2. Parameter Estimation Theory

In general a parameter estimation problem can be stated as follows: Let Z be a random variable having a probability distribution function $F_\theta(Z)$. The subscript in the distribution function indicates that it depends on an unknown parameter θ. In keeping with the standard literature on the subject a canonical personage usually identified as the statistician or decision maker is introduced to facilitate the exposition of the problem. The statistician is interested in estimating the unknown parameter θ from a single observation of the random variable Z. More specifically a function δ, called a decision rule or an estimator, is to be designed so that the random variable $\delta(Z)$ provides the statistician with a "good guess" or estimate of the value of the unknown parameter θ. At this point we must address the following problem: How shall we evaluate the goodness of a given estimate? This is the question of choosing a criterion for estimation. In this work we shall consider the loss function approach to parameter estimation and apply the minimax criterion. This approach and criterion are

Definition 3: Two decision rules are equivalent if they have the same conditional risk function.

Definition 4: A decision rule δ is said to be a Bayes rule if for every other rule γ

$$E\{r[\delta,\theta]\} \leq E\{r[\gamma,\theta]\} \tag{1.2.4}$$

where the average is taken over a prior distribution $\lambda(\theta)$ for θ, that is, δ minimizes the average risk.

Definition 5: δ^* is said to be a minimax decision rule if for all other rules δ

$$\sup_{\theta\in\Omega} r[\delta^*,\theta] \leq \sup_{\theta\in\Omega} r[\delta,\theta]. \tag{1.2.5}$$

That is δ^* infimizes $\sup_{\theta\in\Omega} r[\delta,\theta]$.

1.3 The Relation of Game Theory to Minimax Estimation

Minimax estimation theory is closely related to the theory of two-person zero-sum games. The elements of a two-person zero-sum game are:

(i) The two players, usually identified as the Statistician and Nature.

(ii) The Kernel that represents the pay-offs of the game.

The Kernel is a function of two variables each of which is controlled by one of the players. When a minimax estimation problem is considered as a game, the kernel of the game is identified with the average risk

$$\int_{\Omega} r[\delta,\theta]\, d\lambda(\theta). \tag{1.3.1}$$

The two variables are the decision rule δ controlled by the Statistician and the distribution $\lambda(\theta)$ controlled by Nature.

The rules of the game are such that the Statistician looks for a decision rule δ that minimizes

$$\operatorname*{Sup}_{\lambda} \int_{\Omega} r[\delta,\theta]\, d\lambda(\theta) \tag{1.3.2}$$

while Nature tries to find a distribution $\lambda(\theta)$ that maximizes

$$\inf_{\delta} \int_{\Omega} r[\delta,\theta]\, d\lambda(\theta). \tag{1.3.3}$$

Such a distribution is called a least favorable prior distribution. It can be seen that

$$\operatorname*{Sup}_{\lambda} \int_{\Omega} r[\delta,\theta]\, d\lambda(\theta) = \operatorname*{Sup}_{\theta\in\Omega} r[\delta,\theta], \tag{1.3.4}$$

so that if the Statistician succeeds in the game he has found a minimax solution for the estimation problem.

The fundamental theorem in game theory, the minimax theorem, states that [illegible]r certain conditions

$$\inf_{\delta} \sup_{\lambda} \int_{\Omega} r[\delta,\theta]\, d\lambda(\theta) = \operatorname*{Sup}_{\lambda} \inf_{\delta} \int_{\Omega} r[\delta,\theta]\, d\lambda(\theta). \tag{1.3.5}$$

Now if the minimax theorem holds and a least favorable prior distribution exists, then any minimax rule is Bayes with respect to this prior distribution. See [1] for an extended discussion of Bayesian decision theory and the minimax criterion.

1.4. Preliminary Lemmas and Results

Lemma 1: A Bayes rule is admissible if it is unique up to equivalence.

Proof: (see [1], p. 60).

Lemma 2: (Wald). Let δ_λ be a Bayes rule with respect to $r[\delta,\theta]$ and $\lambda(\theta)$. Assume that $\max\limits_{\theta\in\Omega} r[\delta_\lambda,\theta] = \bar{r}$ exists. Define

$$\tau_\lambda = \{\theta \in \Omega : r[\delta_\lambda,\theta] = \bar{r}\}. \tag{1.4.1}$$

Assume that $\lambda[\tau_\lambda] = 1$. We now have:

(i) δ_λ is minimax

(ii) λ is a least favorable prior distribution.

Proof: (i) (by contradiction). Assume there exists a decision rule δ, and a number $c > 0$ such that

$$\max_{\theta\in\Omega} \{r[\delta,\theta] + c\} < \max_{\theta\in\Omega} r[\delta_\lambda,\theta] \tag{1.4.2}$$

but then

$$\begin{aligned} \int_\Omega \{r[\delta,\theta] + c\} d\lambda(\theta) &\leq \max_{\theta\in\Omega}\{r[\delta,\theta] + c\} \\ &< \max_{\theta\in\Omega} r[\delta_\lambda,\theta] \end{aligned} \tag{1.4.3}$$

and $\max\limits_{\theta\in\Omega} r[\delta_\lambda,\theta] = \bar{r} = \int_\Omega r[\delta_\lambda,\theta]\, d\lambda(\theta)$, since λ assigns probability 1 to τ_λ. Therefore

$$\int_\Omega r[\delta,\theta]\, d\lambda(\theta) + c < \int_\Omega r[\delta_\lambda,\theta] d\lambda \,. \tag{1.4.4}$$

This last inequality contradicts the assumption that δ_λ is a Bayes rule and we therefore conclude that δ_λ is minimax.

(ii) (By contradiction)

Assume there exists another prior probability measure μ and a number $C > 0$ such that

$$\inf_{\delta} \int_{\Omega} r[\delta,\theta]\, d\lambda(\theta) + c < \inf_{\delta} \int_{\Omega} r[\delta,\theta]\, d\mu(\theta) \tag{1.4.5}$$

but

$$\begin{aligned} \int_{\Omega} r[\delta_\lambda,\theta]\, d\mu(\theta) &\geq \inf_{\delta} \int_{\Omega} r[\delta,\theta]\, d\mu(\theta) \\ &\geq \inf_{\delta} \int_{\Omega} r[\delta,\theta]\, d\lambda(\theta) + c = \bar{r} + c \end{aligned} \tag{1.4.6}$$

so that

$$\int_{\Omega} r[\delta_\lambda,\theta]\, d\mu(\theta) \geq \bar{r} + c > \bar{r} \tag{1.4.7}$$

which is a contradiction; therefore λ is a least favorable prior distribution.

2. MINIMAX PARAMETER ESTIMATION UNDER GENERALIZED QUADRATIC LOSS WITH A COMPACT PARAMETER SPACE

2.1. Introduction

The estimation problem we shall consider in this chapter belongs to a general class of problems that can be described by the following model:

Let $z = \theta + v$ where $\theta \in \Omega$, Ω is some specified parameter space contained in E^n, and v is a zero-mean Gaussian random vector with known positive definite covariance matrix Σ. We want to find a minimax estimator for the unknown parameter θ. It is known [2,3,4] that the existence of minimax estimators and the structure of these estimators when they exist is strongly dependent on the structure of the problem which is affected by the underlying loss function, the parameter space Ω and the class of allowed decision rules.

When the loss function is the Standard Quadratic Loss function,

$$L_o[\delta(z),\theta] = [\delta(z)-\theta]'C[\delta(z)-\theta] ,$$

and the parameter space Ω is E^n, a minimax solution for this problem is $\delta(z) = z$; see [1], p. 170 for further details.

The standard quadratic loss function is generalized by adding to it a quadratic term in θ obtaining

$$L_1[\delta(z),\theta] = [\delta(z)-\theta]'C[\delta(z)-\theta] - \theta'D\theta .$$

This loss function forces Nature to compromise between a gain due to increasing the Statistician's estimation error and a loss which is equal to the square of the norm weighted by D. For an unconstrained parameter space, i.e. $\Omega = E^n$, this problem was solved by Basar and Mintz [2].

The parameter θ can be restricted to reside within a known compact subset of E^n, for example a hyperellipsoid defined by

$$\Omega = \{\theta \in E^n : \theta' G\theta \leq e^2, e^2 > 0, G \geq \underline{0}\}.$$

In this case a linear minimax estimator for θ under the standard quadratic loss function was found by Mintz [3]. We will be considering here the problem of finding a linear minimax estimator under a generalized quadratic loss function and will constrain the parameter θ to reside within a known hyperellipsoid in E^n.

2.2. Problem Statement

Let $z = \theta + v$ where v is a normal zero-mean random vector with known positive definite covariance matrix Σ and θ is an unknown element of a known hyperellipsoidal subset of E^n denoted by Ω and defined by

$$\Omega = \{\theta \in E^n : \theta' D_2 \theta \leq e^2; e^2 > 0, D_2 > \underline{0}\} \tag{2.2.1}$$

We wish to determine a minimax estimate $\delta(z)$ for θ with respect to the generalized quadratic loss function

$$L[\delta(z), \theta] = [\delta(z) - \theta]' C[\delta(z) - \theta] - \theta' D_1 \theta \tag{2.2.2}$$

where $C \geq \underline{0}$ and $D_1 > \underline{0}$, subject to the restriction that δ be a linear rule.

2.3. Solution

Define

$$\Gamma(\Lambda,\alpha) \triangleq (\Lambda+\Sigma)^{-1}\Sigma C\Sigma(\Lambda+\Sigma)^{-1} - D_1 - \alpha D_2 \tag{2.3.1}$$

with

$$\Lambda \geq \underline{0} \quad \text{and} \quad \alpha \geq 0. \tag{2.3.2}$$

Let Λ^* and α^* denote any solution to the matrix equation:

$$\Gamma(\Lambda^*,\alpha^*)\Lambda^* = \underline{0} \tag{2.3.3}$$

such that

$$\Gamma(\Lambda^*,\alpha^*) \leq \underline{0} \tag{2.3.4}$$

and

$$\mathrm{Tr}[D_2\Lambda^*] \leq e^2. \tag{2.3.5}$$

Theorem I:

(i) If $\Gamma(\Lambda,\alpha)$ is defined by (2.3.1) and (2.3.2) then there exists a covariance matrix Λ^* and a scalar $\alpha^* \geq 0$ that satisfy (2.3.3) and (2.3.4), and

either (a) $\alpha^* = 0$ and $\mathrm{Tr}[D_2\Lambda^*] = e_*^2 \leq e^2$

or (b) $\alpha^* > 0$ and $\mathrm{Tr}[D_2\Lambda^*] = e_*^2 = e^2$.

(ii) If $\delta(z)$ is defined by

$$\delta^*(z) = \Lambda^*(\Lambda^*+\Sigma)^{-1}z \tag{2.3.6}$$

then $\delta^*(z)$ is a linear minimax estimate for θ. A proof of part (i) of the main theorem may be obtained by generalizing the proof that appears in Appendix I of [3]. The essential modification of that result needed to show part (i) is obtained by replacing $J(A,\Lambda)$ as defined in Lemma 2 of [3] by

$$J(A,\Lambda) \triangleq Tr\{[(I-A)'C(I-A) - D_1]\Lambda\} + Tr[CA\Sigma A']. \tag{2.3.7}$$

We note at this point that the matrix D_2 of this present paper corresponds to D of reference [3]. Since the modified proof is very similar a complete discussion will be omitted. We also note that the last two assertions in Appendix I of [3] do not apply in the present problem due to the difference in the structure of the loss function.

A proof of part (ii) of Theorem I is obtained using Lemma 2 of Section 1.4. We need a prior probability distribution for θ which we define as follows:

Assume $Tr[D_2\Lambda] = e_*^2 \leq e^2$, where $\Lambda \geq \underline{0}$ and consider the matrix $D_2^{\frac{1}{2}}\Lambda D_2^{\frac{1}{2}}$, where $D_2^{\frac{1}{2}}$ is the unique symmetric positive definite square root of D_2. Assume T diagonalizes $D_2^{\frac{1}{2}}\Lambda D_2^{\frac{1}{2}}$, that is

$$S = T'D_2^{\frac{1}{2}}\Lambda D_2^{\frac{1}{2}} T = diag(s_1,\ldots,s_n). \tag{2.3.8}$$

Let m denote an n-dimensional random vector with independent components denoted by m_i , $i=1,\ldots,n$ whose distribution is defined by:

$$Pr\{m_i = s_i^{\frac{1}{2}}\} = \frac{1}{2}, \quad Pr\{m_i = -(s_i^{\frac{1}{2}})\} = \tfrac{1}{2}. \tag{2.3.9}$$

The random vector m takes on 2^k values with equal probability, where k is the rank of Λ. Let $\mu[\Lambda]$ denote the probability measure defined by the distribution of the random vector $y \triangleq D_2^{-\frac{1}{2}}Tm$. We note that the support of the measure $\mu[\Lambda]$ lies entirely in the boundary of Ω^*, a hyperellipsoidal subset of E^n contained in Ω and defined by

$$\Omega^* = \{y \in E^n : y'D_2 y \le e_*^2 \le e^2\}. \tag{2.3.10}$$

This assertion follows from the fact that for each of the 2^k equally likely values of m the vector $y = D_2^{-\frac{1}{2}}Tm$ lies on the boundary of Ω^* since

$$\begin{aligned} y'Dy &= m'T'D_2^{-\frac{1}{2}}D_2 D_2^{-\frac{1}{2}}Tm = m'm \\ &= \sum_{i=1}^{n} s_i = Tr[D_2^{\frac{1}{2}}\Lambda D_2^{\frac{1}{2}}] = Tr[D_2\Lambda] = e_*^2 \le e^2. \end{aligned} \tag{2.3.11}$$

Further, we note that $E[y] = 0$ and

$$\begin{aligned} E[yy'] &= E[D_2^{-\frac{1}{2}}T\, m m'\, T'D_2^{-\frac{1}{2}}] = D_2^{-\frac{1}{2}}TST'D_2^{-\frac{1}{2}} \\ &= D_2^{-\frac{1}{2}}T\,T'\,D_2^{\frac{1}{2}}\Lambda D_2^{\frac{1}{2}}T'\,T\,D_2^{-\frac{1}{2}} \\ &= \Lambda\ . \end{aligned} \tag{2.3.12}$$

Now, let λ_o denote a prior probability distribution for θ such that $\theta \sim \mu[\Lambda]$. For this prior distribution a Bayes estimate is given by

$$\delta_{\lambda_o}(z) = \Lambda(\Lambda + \Sigma)^{-1}z. \tag{2.3.13}$$

This follows from the fact that the loss function is quadratic in the estimation error and therefore within the class of linear rules for a prior distribution with covariance Λ the rule defined by (2.3.13) minimizes the average risk. In what follows we will show that $\lambda^* = \mu[\Lambda^*]$ and $\delta^*(z)$ defined by (2.3.6) satisfy the conditions stated in Lemma 2 Section 1.4

and thus establish part (ii) of the main theorem. We need an expression for the conditional risk function:

$$r[\delta_{\lambda_o},\theta] = E\{L[\delta_{\lambda_o}(z),\theta]|\theta\}$$

$$= \theta'[(\Lambda+\Sigma)^{-1}\Sigma C\Sigma(\Lambda+\Sigma)^{-1} - D_1]\theta$$

$$+ \mathrm{Tr}[(\Lambda+\Sigma)^{-1}\Lambda C\Lambda(\Lambda+\Sigma)^{-1}\Sigma] \ . \tag{2.3.14}$$

The fact that λ^* and $\delta^*(z)$ satisfy Lemma 2 of Section 1.4 is established in several steps as follows:

<u>Step 1</u>: Observe that for all δ of the form $\delta(z) = \Lambda(\Lambda+\Sigma)^{-1}z$. $\Lambda \geq \underline{0}$ the maximum of the conditional risk function

$$\bar{r} = \max_{\theta\in\Omega} r[\delta,\theta]$$

exists since r is continuous in θ and Ω is compact.

<u>Step 2</u>: First consider the case when the solution to (2.3.3) and (2.3.4) satisfy condition (a) of part (i) of the theorem. In this case we note that the conditional risk function can be written as:

$$r[\delta^*,\theta] = \theta'\Gamma(\Lambda^*,0)\theta + \mathrm{Tr}[(\Lambda^*+\Sigma)^{-1}\Lambda^* C\Lambda^*(\Lambda^*+\Sigma)^{-1}\Sigma], \tag{2.3.15}$$

then (2.3.4) implies that τ_{λ^*}, defined by

$$\tau_{\lambda^*} = \{\theta\in\Omega : r[\delta^*,\theta] = \bar{r}\},$$

coincides with the null-space of $\Gamma(\Lambda^*,0)$, that is

$$\tau_{\lambda^*} = N[\Gamma(\Lambda^*,0)]. \tag{2.3.16}$$

When case (b) of part (i) is in force the conditional risk function can be written as

$$r[\delta^*,\theta] = \theta'\Gamma(\Lambda^*,\alpha^*)\theta + \alpha^*\theta'D_2\theta$$

$$+ \mathrm{Tr}[(\Lambda^*+\Sigma)^{-1}\Lambda^* C\Lambda^*(\Lambda^*+\Sigma)^{-1}\Sigma]. \quad (2.3.17)$$

It is obvious that for this case

$$\tau_{\lambda^*} = N[\Gamma(\Lambda^*,\alpha^*)] \cap \partial\Omega \quad (2.3.18)$$

where $\partial\Omega = \{\theta \in E^n : \theta'D_2\theta = e^2\}$. We next prove that in cases (a) and (b) $\lambda^*[\tau_{\lambda^*}] = 1$.

Step 3: In both cases considered in step 2, τ_{λ^*} is a subset of the null-space of $\Gamma(\Lambda^*,\alpha^*)$. We will now prove that λ^* assigns probability one to $N[\Gamma(\Lambda^*,\alpha^*)]$. From (2.3.4) we conclude that $\theta'\Gamma(\Lambda^*,\alpha^*)\theta \leq 0$ but

$$\begin{aligned} E\{\theta'\Gamma(\Lambda^*,\alpha^*)\theta\} &= E\{\mathrm{Tr}[\theta'\Gamma(\Lambda^*,\alpha^*)\theta]\} \\ &= E\{\mathrm{Tr}[\Gamma(\Lambda^*,\alpha^*)\theta\theta']\} \\ &= \mathrm{Tr}[\Gamma(\Lambda^*,\alpha^*)\Lambda^*] \end{aligned} \quad (2.3.19)$$

and (2.3.3) implies that

$$\mathrm{Tr}[\Gamma(\Lambda^*,\alpha^*)\Lambda^*] = 0 . \quad (2.3.20)$$

So we have

$$\theta'\Gamma(\Lambda^*,\alpha^*)\theta \leq 0 \quad (2.3.21a)$$

and

$$E[\theta'\Gamma(\Lambda^*,\alpha^*)\theta] = 0. \quad (2.3.21b)$$

This allows us to conclude that λ^* assigns probability one to the null-space of $\Gamma(\Lambda^*,\alpha^*)$. For the case where Λ^* and α^* satisfy (a) in part (i) this proves that $\lambda^*[\tau_{\lambda^*}] = 1$.

For the situation in which case (b) of part (i) is in force, we note that the discussion following (2.3.10) establishes that $\lambda^* = \mu(\Lambda^*)$ assigns probability one to the boundary of Ω^*, which coincides with Ω in this situation, since $\mathrm{Tr}[D_2\Lambda^*] = e^2$. Hence

$$\lambda^*[\delta\Omega] = 1 \quad \text{and} \quad \lambda^*[N(\Gamma(\Lambda^*,\alpha^*))] = 1$$

and therefore we conclude that τ_{λ^*} as defined by (2.3.18) receives probability one from λ^*. We note at this point that we have a prior probability distribution λ^*, a Bayes estimator $\delta^*(z)$ with respect to λ^* and $r[\delta,\theta]$ and λ^* assigns probability one to the set τ_{λ^*} where $r[\delta^*,\theta]$ attains its maximum value, i.e. $\bar{r}$. Making use of Lemma 2 of Section 1.4 we conclude that

$$\delta^* = \Lambda^*(\Lambda + \Sigma)^{-1}z$$

is a linear minimax estimate for θ. This finishes the proof of part (ii) of the main theorem.

2.4. The Scalar Case

The unidimensional version of the problem stated in Section 2.2 will be treated in this section. This scalar case admits an explicit solution and in obtaining it insight is gained into the structure of the solution to the vector case. We will state the problem and study its solution in what follows. Let $z = \theta + v$ where $v \sim N[0,\sigma^2]$ and θ is an unknown element of Ω defined by $\Omega = \{\theta \in E' : \theta^2 \leq e^2,\ e^2 > 0\}$. We wish to determine a linear minimax estimator $\delta(z)$ with respect to the loss function:

$$L[\delta(z),\theta] = c(\delta - \theta)^2 - d\theta^2 \tag{2.4.1}$$

where $c > 0$ and $d > 0$. For this loss function the conditional risk function for linear decision rules of the form $\delta = az$ is given by

$$r[\delta,\theta] = [c(a-1)^2 - d]\theta^2 + ca^2\sigma^2. \tag{2.4.2}$$

There are two cases to consider depending on the relative magnitudes of c and d.

Case 1: $c-d < 0$

For this case an appropriate prior distribution for θ is $\lambda^*(\theta)$ such that $\theta = 0$ with probability one. The average conditional risk with respect to λ^* is

$$E_{\lambda^*}\{r[\delta,\theta]\} = ca^2\sigma^2. \tag{2.4.3}$$

A Bayes rule is one that minimizes (2.4.3) which turns to be positive except when $a = 0$ so that

$$\delta^*(z) = 0 \tag{2.4.4}$$

is the required Bayes rule. Using the rule defined by (2.4.4) in (2.4.2) the conditional risk function results

$$r[\delta,\theta] = [c - d]\theta^2 \tag{2.4.5}$$

since $c-d < 0$ $\bar{r} = 0$ occurs when $\theta = 0$. This is the point that receives probability one from λ^* and therefore $\delta^*(z) = 0$ is minimax according to Lemma 2 in Section 1.4.

Case 2: $c-d \geq 0$

If

$$e^2 \geq \sigma^2(\sqrt{\frac{c}{d}} - 1), \tag{2.4.6}$$

then λ^* assigning equal probability to

$$\theta = \pm q = \pm \sigma(\sqrt{\frac{c}{d}} - 1)^{\frac{1}{2}} \tag{2.4.7}$$

is an appropriate prior distribution. With this prior distribution the average conditional risk is

$$E_{\lambda^*}\{r[\delta,\theta]\} = [c(a-1)^2 - d]q^2 + ca^2\sigma^2. \tag{2.4.8}$$

By elementary calculus the value of a that minimizes (2.4.8) is

$$a^* = \frac{q^2}{q^2 + \sigma^2} \tag{2.4.9}$$

and the corresponding Bayes rule results

$$\delta^* = a^*z. \tag{2.4.10}$$

Use of this rule in the conditional risk yields a constant value independent of the value of θ in Ω, i.e.,

$$r[\delta^*,\theta] = c(a^*)^2\sigma^2.$$

This particular situation characterizes δ^* as an equalizer rule, i.e. one that produces a constant conditional risk. In this case τ_{λ^*} is identified with the whole parameter space Ω, the condition of Lemma 2 Section 1.4 that $\lambda^*[\tau_{\lambda^*}]$ be one is thus trivially satisfied. We note that all the conditions of Lemma 2 have been established and we conclude that in this case 2 under the condition defined in (2.4.6) the decision rule defined by (2.4.7), (2.4.9), and (2.4.10) in a minimax rule.

On the other hand, if condition (2.4.6) is not satisfied but

$$e^2 < \sigma^2(\sqrt{\frac{c}{d}} - 1) \qquad (2.4.11)$$

we define a prior probability distribution for θ by letting λ^* assign equal probabilities to $\theta = \pm e$, that is:

$$\Pr\{\theta = e\} = \Pr\{\theta = -e\} = \frac{1}{2} . \qquad (2.4.13)$$

With this probability distribution the average conditional risk becomes

$$E_{\lambda^*}\{r[\delta,\theta]\} = [c(a-1)^2 - d]e^2 + ca^2\sigma^2. \qquad (2.4.13)$$

By reasoning similar to that which lead to (2.4.9), the value of the multiplier a^* defining the Bayes rule for this case is

$$a^* = \frac{e^2}{e^2 + \sigma^2} , \qquad (2.4.14)$$

and the Bayes rule is

$$\delta^* = a^* z . \qquad (2.4.15)$$

The conditional risk function for this Bayes rule is

$$r[\delta^*,\theta] = [c(a^*-1)^2 - d]\theta^2 + c(a^*)^2\sigma^2. \tag{2.4.16}$$

The conditional risk is quadratic in θ and the coefficient of the quadratic term is positive. This assertion follows from equation (2.4.14) defining a^* and from condition (2.4.11). The maximum value of the conditional risk occurs then at the extreme points of the parameter interval, i.e.

$$\tau_{\lambda^*} = \{-e,e\}. \tag{2.4.17}$$

Obviously $\lambda^*[\tau_{\lambda^*}] = 1$, and according to Lemma 2 of Section 1.4

$$\delta^* = \frac{e^2}{e^2+\sigma^2}\, z$$

is a linear minimax estimate for θ.

2.5. Remarks

The main theorem in Section 2.3 provides us with a linear minimax estimate of the unknown parameter θ with respect to the generalized quadratic loss function defined in Section 2.2 where the parameter space is defined to be a hyperellipsoid Ω subset of E^n. The least favorable prior distribution which we associated with this minimax estimate is a discrete probability distribution that has support on the boundary of a hyperellipsoidal subset of E^n, which we denoted by Ω^*, contained in the original hyperellipsoid Ω. The minimax estimator, the least favorable prior distribution, and the hyperellipsoid Ω^* are all determined by the solution to the algebraic matrix equations stated in part (i) of the theorem of Section 2.3.

When the restriction imposed on the parameter that it reside in this hyperellipsoid is lifted, Basar and Mintz [2] showed that the minimax estimate, over linear and nonlinear rules, is a linear estimate and the least favorable prior distribution is a zero-mean normal distribution. A closely related problem solved by Mintz [3] in which the loss function is the standard quadratic loss function and the parameter space is a hyper-ellipsoidal subset of E^n exhibits a similar structure in the least favorable prior distribution to the problem treated here, with the difference that in that case the support of the discrete prior distribution lies always on the boundary of the parameter space. The class of decision rules considered in [3] is also restricted to linear rules.

2.6. A Generalized Model

We consider now a more general observation model. Let $z = x + \theta + v$, where x and v are independent zero mean Gaussian random vectors with positive definite covariance matrices Q and Σ respectively. The results obtained in Section 2.3 can be extended to find a linear minimax estimate for $u \triangleq x + \theta$ under a generalized quadratic loss function given by

$$L[\delta(z),u] = (\delta-u)'c(\delta-u) - \theta'D_1\theta.$$

The unknown parameter θ is an element of $\Omega = \{\theta \in E^n : \theta'D_2\theta \le e^2,\ e^2 > 0\}$. The following theorem provides the desired minimax estimate.

Theorem II:

Define

$$\Gamma(\Lambda,\alpha) \triangleq (\Lambda + Q + \Sigma)^{-1}\Sigma C\Sigma(\Lambda + Q + \Sigma)^{-1} - D_1 \cdot \alpha D_2. \qquad (2.6.1)$$

Let Λ^* and α^* denote any solution to

$$\Gamma(\Lambda^*,\alpha^*)\Lambda^* = 0 \tag{2.6.2}$$

$$\Gamma(\Lambda^*,\alpha^*) \leq 0 \tag{2.6.3}$$

$$\mathrm{Tr}(D_2\Lambda^*) \leq e^2 \tag{2.6.4}$$

where

$$\Lambda^* \geq \underline{0} \quad \text{and} \quad \alpha^* \geq 0. \tag{2.6.5}$$

With this definition and notation we conclude that:

(i) There exists a pair Λ^* and α^* satisfying (2.6.2) through (2.6.5), and either

$$\alpha^* = 0 \quad \text{and} \quad \mathrm{Tr}[D_2\Lambda^*] = e_*^2 \leq e^2$$

or

$$\alpha^* > 0 \quad \text{and} \quad \mathrm{Tr}[D_2\Lambda^*] = e_*^2 = e^2.$$

(ii) The decision rule δ^* given by

$$\delta^*(z) = (\Lambda^* + Q)(\Lambda^* + Q + \Sigma)^{-1}z \tag{2.6.6}$$

is a linear minimax estimate for u. The details of the proof are omitted here since it is essentially the same as the proof of Theorem 1 of Section 2.3.

A further extension of the work presented in this thesis can be used to determine a linear minimax terminal state estimate for a linear plant subject to an energy constraint with a generalized quadratic loss function as the underlying loss function.

REFERENCES

1. T. S. Ferguson, Mathematical Statistics - A Decision Theoretic Approach, Academic Press, New York, N. Y., 1967.

2. T. Basar and M. Mintz, "On a Minimax Estimate for the Mean of a Normal Random Vector Under a Generalized Quadratic Loss Function," The Annals of Statistics, Vol. 1, No. 1, pp. 127-134, 1973.

3. M. Mintz, "On a Minimax Parameter Estimation Problem with a Compact Parameter Space," Proceedings of the 1972 IEEE Conference on Decision and Control, pp. 178-182.

4. D. J. Stelte and M. Mintz, "On Minimax Parameter Estimation under Exponentiated Generalized Quadratic Loss," Proceedings of the Eleventh Allerton Conference on Circuit and System Theory, Univ. of Ill., pp. 215-225, October 1973.